(19) JAPANESE PATENT OFFICE

PATENT ABSTRACTS OF JAPAN

(11) Publication number: 06149482 A

(43) Date of publication of application: 27.05.94

(51) Int. Cl G06F 3/06

(21) Application number: 04301367

(22) Date of filing: 11.11.92

(71) Applicant: HITACHI LTD

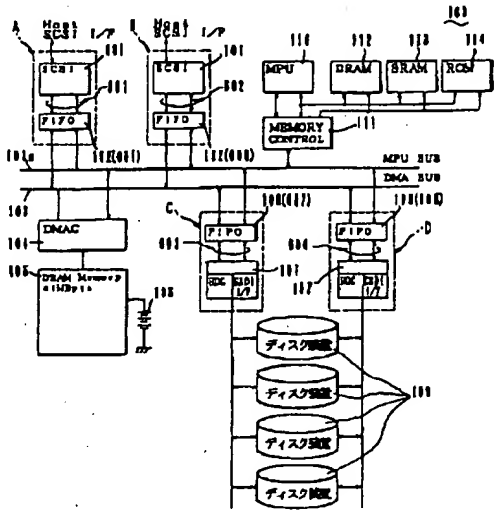
(72) Inventor: SATO MASAHIKO
NAKAKOSHI KAZUO
TAKAHASHI NAOYA
CHUMA AKIRA
YUGAWA YOSHIO

(54) EXTERNAL STORAGE DEVICE

(57) Abstract:

PURPOSE: To provide an external storage device capable of shortening time for responding to respective read/write requests from a host device and substantially increasing the issuing number of the read/write requests from the host device.

CONSTITUTION: In a controller part 100 interposed between disk devices 109 and the host device, the buffer memory 105 of a large capacity nonvolatilized by a battery 106, plural host interfaces A and B, plural drive interfaces C and D, a DMA bus 103 and a DMA controller 104 are provided. By the reconnection by dynamic selection and the release of the host interfaces A and B and the drive interfaces C and D, a read/write processing from the host device to a buffer memory 105 and the read/write processing between the buffer memory 105 and the disk devices 109 are parallelly and asynchronously performed.



(19)日本国特許庁（ＪＰ）　(12)公開特許公報（Ａ）　(11)特許出願公開番号

特開平6－149482

(43)公開日　平成６年(1994)５月27日

(51)Int.Cl.[5]	識別記号	庁内整理番号	ＦＩ	技術表示箇所
Ｇ０６Ｆ　3/06	３０２　Ａ	7165－5B		

審査請求　未請求　請求項の数８(全 13 頁)

(21)出願番号　特願平4－301367

(22)出願日　平成４年(1992)11月11日

(71)出願人　000005108
株式会社日立製作所
東京都千代田区神田駿河台四丁目６番地

(72)発明者　佐藤　雅彦
神奈川県小田原市国府津2880番地　株式会社日立製作所ストレージシステム事業部内

(72)発明者　中越　和夫
神奈川県小田原市国府津2880番地　株式会社日立製作所ストレージシステム事業部内

(72)発明者　高橋　直也
神奈川県小田原市国府津2880番地　株式会社日立製作所ストレージシステム事業部内

(74)代理人　弁理士　筒井　大和

最終頁に続く

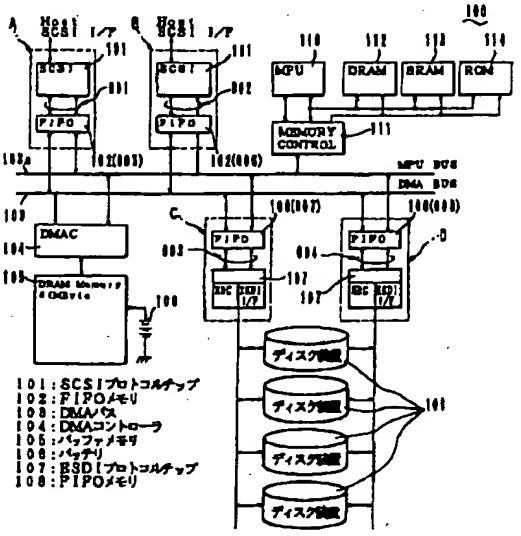
(54)【発明の名称】　外部記憶装置

(57)【要約】

【目的】　上位装置からの個々のリード／ライト要求に対する応答時間の短縮、および上位装置からのリード／ライト要求の発行件数の大幅な増加を実現することが可能な外部記憶装置を提供する。

【構成】　ディスク装置１０９と上位装置との間に介在するコントローラ部１００に、バッテリ１０６によって不揮発化された大容量のバッファメモリ１０５と、複数口のホストインターフェイスA，Bおよび複数口のドライブインターフェイスC，Dと、DMAバス１０３およびDMAコントローラ１０４を設け、ホストインターフェイスA，BおよびドライブインターフェイスB，Cの解放および動的な選択による再接続により、上位装置からバッファメモリ１０５に対するリード／ライト処理と、バッファメモリ１０５とディスク装置１０９との間のリード／ライト処理を非同期に並行して行うようにした外部記憶装置である。

図 1



1

【特許請求の範囲】

【請求項1】 データを第1の記憶媒体に持久的に記憶するドライブと、このドライブと上位装置との間における前記データの授受を制御する入出力制御部とを備えた外部記憶装置であって、前記入出力制御部は、前記第1の記憶媒体よりも前記データに対するアクセスをより高速に行うことが可能な第2の記憶媒体からなる一時データ保持手段、および前記上位装置からの前記ドライブに対する前記データの書き込み要求に対し、当該書き込みデータを前記一時データ保持手段に一旦書き込み、当該書き込み動作が完了した時点で前記上位装置に対して書き込み完了を応答する手段、および前記一時データ保持手段から前記ドライブへの前記データの書き込みを、前記上位装置からの前記データの書き込みおよび読み出しのタイミングとは独立に前記入出力制御部で管理されたタイミングで行う手段を含むことを特徴とする外部記憶装置。

【請求項2】 前記入出力制御部は、前記上位装置からの前記データの読み出し要求に対し、前記ドライブから読み出したデータを前記一時データ保持手段に残す手段、および前記上位装置からの同一の前記データの読み出し要求に対し、前記一時データ保持手段に残された前記データを読み出して転送する手段を含むことを特徴とする請求項1記載の外部記憶装置。

【請求項3】 前記第1の記憶媒体が磁気ディスクであり、前記第2の記憶媒体が半導体メモリからなり、当該半導体メモリはセルフ・リフレッシュ機能を有するダイナミック・ランダム・アクセス・メモリ（ＤＲＡＭ）で構成され、前記半導体メモリの電源は、前記入出力制御部と共通の主電源と、この主電源とは独立なバッテリの2系統で構成され、前記バッテリの充電が前記主電源によって行われるようにしたことを特徴とする請求項1または2記載の外部記憶装置。

【請求項4】 前記上位装置との間の第1のインターフェイスポートを複数口有し、前記入出力制御部は、前記上位装置から発行される前記データの読み出しまたは書き込みコマンドに対し、当該データの読み出しまたは書き込み準備が完了するまでの間、前記第1のインターフェイスポートを解放する手段、および読み出しまたは書き込みの準備が完了した時点で、前記第1のインターフェイスポートを再接続する手段、および当該再接続に際して、複数の前記第1のインターフェイスポートのうち、空いているポートを選択し、動的に第1のインターフェイスポートの接続を行う手段、および前記動的な第1のインターフェイスポートの接続を抑止する手段を含むことを特徴とする請求項1，2または3記載の外部記憶装置。

【請求項5】 前記入出力制御部に対して前記ドライブをディージーチェーンで複数台接続するとともに、当該入出力制御部と当該ドライブとの間に第2のインターフェイスポートを複数口設け、前記入出力制御部は、この第2のインターフェイスポートによる前記ドライブに対する前記データの読み出しまたは書き込みにおいて、前記ドライブにおける前記データの読み出しまたは書き込み準備が完了するまでの間、当該第2のインターフェイスポートを解放する手段、および前記ドライブにおける前記データの読み出しまたは書き込み準備が完了した時点で、当該ドライブから入力される接続要求割り込みに応じて当該第2のインターフェイスポートを再接続する手段、および当該第2のインターフェイスポートの再接続に際して、複数の前記第2のインターフェイスポートのうち、空いている側を選択して動的に再接続する手段を含むことを特徴とする請求項1，2，3または4記載の外部記憶装置。

2

【請求項6】 前記入出力制御部は、全体の制御を司るマイクロプロセッサと、複数の前記第1のインターフェイスポートおよび前記第2のインターフェイスポートと前記一時データ保持手段とを時分割処理によって切替えながら接続することによりＤＭＡ（ダイレクト・メモリ・アクセス）転送を行うＤＭＡコントローラと、前記マイクロプロセッサと前記第1，第2のインターフェイスポートおよび前記ＤＭＡコントローラを接続するＭＰＵバスと、前記ＤＭＡコントローラと前記第1，第2のインターフェイスポートを接続するＤＭＡバスと、前記第1および第2のインターフェイスポートと前記ＭＰＵバスおよびＤＭＡバスとの間に設けられ、前記一時データ保持手段といずれかのインターフェイスポートとの間でＤＭＡ転送が行われている間、他のインターフェイスポートを停止させないためのＦＩＦＯ（先入れ／先出し）メモリとを含むことを特徴とする請求項1，2，3，4または5記載の外部記憶装置。

【請求項7】 ＤＭＡバスのバス幅を、前記第1および第2のインターフェイスポートのバス幅よりも大きくしたことを特徴とする請求項1，2，3，4，5または6記載の外部記憶装置。

【請求項8】 前記ＤＭＡバスと、前記第1および第2のインターフェイスポートのバス幅の変換を、前記ＦＩＦＯメモリの入出力バス幅を当該ＤＭＡバス側と当該第1および第2のインターフェイスポート側で変えることにより実現することを特徴とする請求項1，2，3，4，5，6または7記載の外部記憶装置。

【発明の詳細な説明】

【０００１】

【産業上の利用分野】本発明は外部記憶装置に関し、特に、ＯＬＴＰ（Online Transaction Process），ＲＤＢ（Relational Data Base）等、高スループット，高速Ｉ／Ｏを要求される情報処理システムに使用される磁気ディスク装置などの外部記憶装置に適用して有効な技術に関する。

【０００２】

【従来の技術】高性能のマイクロプロセッサの登場により、計算機システムのデータ処理能力は、急速に高速化され、また、データベース等の大容量化に伴い、これら計算機システムで扱われるデータの規模も大型化の傾向にある。

【０００３】このような背景のもと、これらの計算機システムから外部記憶装置の一種である磁気ディスク装置に発行される単位時間当たりのＩ／Ｏ要求件数は、近年急速に大きくなってきている。しかし、磁気ディスク装置は、その動作原理上、データアクセス時に機械的な動作が伴うため、そのアクセス速度は、計算機のＣＰＵな主メモリの速度と比べ大きな隔たりを持っている。このため、磁気ディスク装置が計算機システムの処理能力を決定する上でのボトルネックとなり、計算機システムの性能が磁気ディスク装置のスループットに大きく影響されるといった傾向がある。

【０００４】従来、このような問題を改善する手段としてＬＲＵ (Least Recently Used)方式を用いたリードキャッシュを用い、記憶媒体上において時間的、物理的近傍でアクセスされたデータはキャッシュメモリに残しておき、上位装置からのリード要求がキャッシュメモリ内のデータであった場合（キャッシュヒット）には、磁気ディスクに対してアクセスすることなくキャッシュメモリ上のデータを転送し、応答時間の短縮を図ることが行われていた。また、特開昭６０－７４５７号公報に開示される技術では、複数台の磁気ディスク装置に複数台の磁気ディスクコントローラを接続し、デバイスクロスコール方式を改善した技術が提示されているが、この技術も同様にＬＲＵ方式を用いたリードキャッシュ方式の応用にしかすぎず、データのライト処理に対しては磁気ディスク装置の性能の低さがそのままシステムの処理能力に反映されてしまう、といった不具合があった。

【０００５】また、従来の磁気ディスクコントローラは、上位装置からのリードおよびライトコマンドに対し、下位ディスク装置にシークコマンドを発行し、ヘッドが目的のトラックに到達するまでのシーク時間、および目的のトラック上のセクタがヘッド下に到来するまでの回転待ち時間の間、インターフェイスを接続状態のままとしており、上位装置から極めて高頻度のＩ／Ｏ要求が発行された場合、インターフェイスのバスネックとなり、Ｉ／Ｏ発行件数が制約されるという問題があった。

【０００６】

【発明が解決しようとする課題】磁気ディスク装置に対する高速化の要求として、Ｉ／Ｏ性能の向上、すなわち、高速にＩ／Ｏ処理を実行すること、および、単位時間当たりのＩ／Ｏ発行件数の向上がある。

【０００７】上記従来技術では、いずれの場合もデータの読み出し処理についてはそれなりの効果があるものの、データの書き込みについては配慮されておらず、磁気ディスク装置のトータルのＩ／Ｏ性能としては、必ずしも計算機システムの要求性能を満足するものではなかった。

【０００８】また、近年、計算機システムの外部記憶装置に要求される記憶容量の増大、およびクライアント／サーバ方式などに代表される分散処理の流れから、磁気ディスクサブシステムに複数台の磁気ディスクドライブを内蔵することで、大容量化を実現し、このような構成のサブシステムに対して、複数台のホストコンピュータからのアクセス要求が発行される様な環境下で磁気ディスクサブシステムが稼動することが現実となってきている。

【０００９】このような状況から、磁気ディスクサブシステム内のコントローラは、複数台のホストコンピュータと、複数台の磁気ディスクドライブの接続を制御しなくてはならなくなって来ている。

【００１０】上記従来技術では、ホストインターフェイスおよびドライブインターフェイスは、それぞれディージーチェーン方式で複数台のホストコンピュータおよび複数台の磁気ディスクドライブに接続され、特定のホストコンピュータからのＩ／Ｏ要求が発生した場合には、当該Ｉ／Ｏ要求が終了するまで、それぞれのホストコンピュータは待ち状態となり、インターフェイスのバスネックによりＩ／Ｏ発行件数が制限されるといった問題があった。この対策として、たとえばインターフェイスバスを個別に持たせることも考えられるが、コントローラが非常に高価かつ大規模になってしまうという他の問題を生じてしまう。

【００１１】本発明の目的は、上位装置からの個々のリード／ライト要求に対する応答時間の短縮、および上位装置からのリード／ライト要求の発行件数の大幅な増加を実現することが可能な外部記憶装置を提供することにある。

【００１２】本発明の前記ならびにその他の目的と新規な特徴は、本明細書の記述および添付図面から明らかになるであろう。

【００１３】

【課題を解決するための手段】本願において開示される発明のうち、代表的なものの概要を簡単に説明すれば、下記のとおりである。

【００１４】すなわち、本発明は、データを第１の記憶媒体に持久的に記憶するドライブと、このドライブと上位装置との間におけるデータの授受を制御する入出力制御部とを備えた外部記憶装置において、入出力制御部は、第１の記憶媒体よりもデータに対するアクセスをより高速に行うことが可能な第２の記憶媒体からなる一時データ保持手段、および上位装置からのドライブに対するデータの書き込み要求に対し、当該書き込みデータを一時データ保持手段に一旦書き込み、当該書き込み動作が完了した時点で上位装置に対して書き込み完了を応答する手段、および一時データ保持手段からドライブへの

データの書き込みを、上位装置からのデータの書き込みおよび読み出しのタイミングとは独立に入出力制御部で管理されたタイミングで行う手段、を有したものである。

【００１５】また、本発明は、請求項１記載の外部記憶装置において、入出力制御部には、上位装置からのデータの読み出し要求に対し、ドライブから読み出したデータを一時データ保持手段に残す手段、および上位装置からの同一のデータの読み出し要求に対し、一時データ保持手段に残されたデータを読み出して転送する手段を設けたものである。

【００１６】また、本発明は、請求項１または２記載の外部記憶装置において、第１の記憶媒体が磁気ディスクであり、第２の記憶媒体が半導体メモリからなり、当該半導体メモリはセルフ・リフレッシュ機能を有するダイナミック・ランダム・アクセス・メモリ（ＤＲＡＭ）で構成され、半導体メモリの電源は、入出力制御部と共通の主電源と、この主電源とは独立なバッテリの２系統で構成され、バッテリの充電が主電源によって行われるようにしたものである。

【００１７】また、本発明は、請求項１，２または３記載の外部記憶装置において、上位装置との間の第１のインターフェイスポートを複数口有し、入出力制御部には、上位装置から発行されるデータの読み出しまたは書き込みコマンドに対し、当該データの読み出しまたは書き込み準備が完了するまでの間、第１のインターフェイスポートを解放する手段、および読み出しまたは書き込みの準備が完了した時点で、第１のインターフェイスポートを再接続する手段、および当該再接続に際して、複数の第１のインターフェイスポートのうち、空いているポートを選択し、動的に第１のインターフェイスポートの接続を行う手段、および動的な第１のインターフェイスポートの接続を抑止する手段を設けたものである。

【００１８】また、本発明は、請求項１，２，３または４記載の外部記憶装置において、入出力制御部に対してドライブをディージーチェーンで複数台接続するとともに、当該入出力制御部と当該ドライブとの間に第２のインターフェイスポートを複数口設け、入出力制御部には、この第２のインターフェイスポートによるドライブに対するデータの読み出しまたは書き込みにおいて、ドライブにおけるデータの読み出しまたは書き込み準備が完了するまでの間、当該第２のインターフェイスポートを解放する手段、およびドライブにおけるデータの読み出しまたは書き込み準備が完了した時点で、当該ドライブから入力される接続要求割り込みに応じて当該第２のインターフェイスポートを再接続する手段、および当該第２のインターフェイスポートの再接続に際して、複数の第２のインターフェイスポートのうち、空いている側を選択して動的に再接続する手段を備えたものである。

【００１９】また、本発明は、請求項１，２，３，４または５記載の外部記憶装置において、入出力制御部は、全体の制御を司るマイクロプロセッサと、複数の第１のインターフェイスポートおよび第２のインターフェイスポートと一時データ保持手段とを時分割処理によって切替えながら接続することによりＤＭＡ（ダイレクト・メモリ・アクセス）転送を行うＤＭＡコントローラと、マイクロプロセッサと第１，第２のインターフェイスポートおよびＤＭＡコントローラを接続するＭＰＵバスと、ＤＭＡコントローラと第１，第２のインターフェイスポートを接続するＤＭＡバスと、第１および第２のインターフェイスポートとＭＰＵバスおよびＤＭＡバスとの間に設けられ、一時データ保持手段といずれかのインターフェイスポートとの間でＤＭＡ転送が行われている間、他のインターフェイスポートを停止させないためのＦＩＦＯ（先入れ／先出し）メモリとを含む構成としたものである。

【００２０】また、本発明は、請求項１，２，３，４，５または６記載の外部記憶装置において、ＤＭＡバスのバス幅を、第１および第２のインターフェイスポートのバス幅よりも大きくしたものである。

【００２１】また、本発明は、請求項１，２，３，４，５，６または７記載の外部記憶装置において、ＤＭＡバスと、第１および第２のインターフェイスポートのバス幅の変換を、ＦＩＦＯメモリの入出力バス幅を当該ＤＭＡバス側と当該第１および第２のインターフェイスポート側で変えることにより実現するものである。

【００２２】

【作用】通常、たとえば磁気ディスク装置などの外部記憶装置においては、データの書き込みは、ヘッドを目的のシリンダに移動し、さらにディスクの回転によってヘッドが目的のセクタ上に到達した後に行われる。この場合、ヘッドを移動するシーク時間、およびヘッドの直下に目的のセクタが到来するまでの回転待ち時間などの機械的な動作時間がオーバーヘッドとして加算され、書き込み時のＩ／Ｏ性能を制限している。

【００２３】上記した本発明の外部記憶装置では、ドライブの第１の記憶媒体よりも遥に高速な半導体メモリなどからなる一時データ保持手段に対して、リードデータのみならず、書き込み要求に際してはライトデータも格納し、当該一時データ保持手段に対するライトデータの書き込みが完了した時点で、上位装置に対して書き込み動作の完了を報告し、一時データ保持手段からドライブへの実際の書き込みは、上位装置からの書き込み要求の契機とは非同期に実行するので、ドライブ側における前述のような機械的な要因の各種待ち時間が書き込みデータの処理の所要時間に影響することがなくなり、上位装置から発行される個々のＩ／Ｏ要求をドライブの動作性能に無関係に高速に処理することができる。

【００２４】また、下位のドライブとの間の第２のインターフェイスポートにおいて、ドライブに対してコマン

ドを発行した時点で当該ポートを解放し、ドライブ側の準備が完了した時点で再接続の割り込み要求を発行して当該ポートを再接続する動作を行うので、あるドライブの準備の間は第2のインターフェイスポートを他のドライブとのI／O処理に用いることが可能となる。また、さらに、第2のインターフェイスポートを複数備え、前記再接続などに際しては空いているインターフェイスポートを動的に選択して接続することにより、複数の第2のインターフェイスポートを無駄なく使用でき、しかも再接続の待ち時間も短縮される。この結果、単位時間当たりに受付処理可能なI／Oの発行件数が増加し、複数の上位装置によって共有される場合に発生するデバイスクロスコールなどに的確に対応することができる。

【0025】

【実施例】以下、本発明の一実施例である外部記憶装置を図面を参照しながら詳細に説明する。

【0026】図1は本実施例の外部記憶装置の構成の一例を示すブロック図である。本実施例では、外部記憶装置の一例として、磁気ディスクサブシステムの場合について説明する。

【0027】本実施例の磁気ディスクサブシステムは、たとえば磁気ディスクなどを記憶媒体としてデータを持久的に記憶する複数台のディスク装置109と、このディスク装置109と図示しないホストコンピュータ（上位装置）（以下、ホストと略記する）との間におけるデータの授受を制御するコントローラ部100とを含んでいる。

【0028】コントローラ部100のMPUバス103aおよびDMAバス103には、図示しないホストとの間に介設される2系統のホストインターフェイスA，ホストインターフェイスB、および複数台のディスク装置109がディジーチェーン接続される2系統のドライブインターフェイスCおよびドライブインターフェイスDが接続されている。

【0029】複数のホストインターフェイスA，Bの各々は、ホストとの間でSCSI (Small Computer System Interface)プロトコルによる情報の授受を制御するSCSIプロトコルチップ101と、このSCSIプロトコルチップ101とMPUバス103aおよびDMAバス103との間に介在するFIFOメモリ102によって構成されている。

【0030】一方、複数のドライブインターフェイスC，Dの各々は、コントローラ部100に接続される複数のディスク装置109との間において、ESDI (EnhancedSmall Device Interface)プロトコルによって情報の授受を行うESDIプロトコルチップ107と、このESDIプロトコルチップ107とMPUバス103aおよびDMAバス103との間に介在するFIFOメモリ108によって構成されている。

【0031】この場合、MPUバス103aおよびDMAバス103には、DMAコントローラ104を介してバッファメモリ105が接続されている。このバッファメモリ105は、たとえば容量が64メガバイトの半導体メモリ（セルフ・リフレッシュ機能付のDRAM）で構成されている。さらに、バッファメモリ105は図示しない主電源と、当該主電源によって常時充電されているバッテリ106の双方から給電される構成となっており、主電源が切断された場合、バッテリ106からの給電によって、たとえば最長一週間、記憶データを保持することが可能になっている。

【0032】また、MPUバス103aには、バスコントローラ111を介して、本実施例の磁気ディスクサブシステムの全体の制御を行うマイクロプロセッサ110（MPU），当該マイクロプロセッサ110の制御プログラムやワーク用メモリエリアを提供する制御メモリ112（DRAM），制御メモリ113（SRAM），制御メモリ114（ROM）が接続されており、後述のような一連の制御動作が実現される構成となっている。

【0033】以下、本実施例の磁気ディスクサブシステムの作用の一例について説明する。

【0034】図2および図3に、図示しない上位のホストからディスク装置109に対してデータを書き込む場合のデータの流れを示す。

【0035】図2は、ホストからバッファメモリ105にデータを転送するまでの処理を示している。ホストからデータの書き込み要求が発生した場合、SCSIプロトコルチップ101から割り込み信号201が発生し、マイクロプロセッサ110はデータの書き込み要求を認識する。マイクロプロセッサ110は、この要求に従い、バスコントローラ111およびMPUバス103aを介して制御アクセス202，制御アクセス203，制御アクセス204を行うことにより、SCSIプロトコルチップ101，FIFOメモリ102，DMAコントローラ104に対してDMA転送のための各種レジスタの設定などを行う。

【0036】以上の設定が終了すると、ホストからのデータは、バス206でバッファメモリ105に対してDMA転送される。また、このDMA転送の間、マイクロプロセッサ110は、制御アクセス205により、ESDIプロトコルチップ107を介してディスク装置109に対して、データを格納すべきシリンダへのシーク命令の発行およびヘッドの切替えなどを行う。

【0037】ホストからのDMA転送と、ディスク装置109におけるシークが終了した時点でマイクロプロセッサ110は、次に、バッファメモリ105に格納された書き込みデータ207をディスク装置109に転送する。

【0038】この時の書き込みデータ207の流れを図3に示す。マイクロプロセッサ110は、制御アクセス301，制御アクセス302，制御アクセス303によ

り、DMAコントローラ１０４，ＥＳＤＩプロトコルチップ１０７，ＦＩＦＯメモリ１０８に対して、DMA転送のための設定を行う。以上の設定が終了すると、バッファメモリ１０５上の書き込みデータ２０７は、バス３０４でバッファメモリ１０５からディスク装置１０９に転送され、目的のディスク装置１０９の格納位置３０５に書き込まれる。

【００３９】以上の制御を図４のタイムチャートで示す。同図では、ＳＣＳＩプロトコルチップ１０１，DMAバス１０３，ディスク装置１０９（Ｄｒｉｖｅ）の各々における処理動作を並べて示している。書き込み要求が発生すると、ホストは、コントローラ部１００に対して区間４０１の間にデータ転送を行う。これと同時に、コントローラ部１００の内部でＳＣＳＩプロトコルチップ１０１からバッファメモリ１０５に対してDMA転送が開始され、区間４０２でデータ転送が行われる。DMA転送が若干遅れるのは、ＦＩＦＯメモリ１０２でデータをバッファリングしているためで、この遅れ時間は、転送所要時間に比べて極めて短い時間である。また、前記転送動作と同時にディスク装置１０９に対してシークコマンドが発行され、ディスク装置１０９は、区間４０３でシークを実行し、シークが完了するとさらに区間４０４で回転待ちを行い、データを書き込むべきセクタがヘッドの直下に到来すると区間４０５でバッファメモリ１０５からディスク装置１０９にDMA転送が開始され、区間４０６でディスク装置１０９に対するデータの書き込みが実行される。

【００４０】従来のディスクコントローラでは、ホストからデータを受理し、ディスク装置１０９に対するデータの書き込みが完了した区間４０７でコマンド終了報告を応答していたため、転送開始から区間４０７までの区間４０８が一回の書き込み動作の所要時間となっていた。

【００４１】これに対して、本実施例の場合には、区間４０２のDMA転送が終了し、バッファメモリ１０５に対する書き込みデータの転送完了の区間４０９でホストに対して書き込みコマンドの終了報告を行っている。このため、ホストから見れば、ディスク装置１０９におけるシークの区間４０３と回転待ちの区間４０４を省いた区間４１０が一回の書き込みコマンドの所要時間となり、従来に比較して、極めて高速にコマンドが実行されたことになる。ディスク装置１０９への書き込みは、ホストに対する応答処理とは非同期に実行される。すなわち、ホストからのＩ／Ｏ要求が混み合っている場合は、ディスク装置１０９へのデータの書き込みは一時保留し、Ｉ／Ｏ要求頻度が低下した時点で実行することでさらにホストからのコマンドを高速に処理することができる。

【００４２】バッファメモリ１０５に空き領域が無くなった場合は、未反映のデータのディスク装置１０９への書き込み終了を待ってから、ホストからのデータの受領を行うため、従来と同程度のスループットに低下するが、本実施例のように、バッファメモリ１０５の容量を大きく取った場合、当該事象は極めて多くのＩ／Ｏ要求が発行されるまで発生することはない。

【００４３】ここで問題となることは、ホストからのバッファメモリ１０５への書き込みデータのDMA転送が終了し、ホストにコマンド終了報告を行った後、データがディスク装置１０９に実際に書き込まれる前に、なんらかの原因で主電源が切断された場合のデータ喪失の危険性である。

【００４４】本実施例の場合、前述のようにバッファメモリ１０５をバッテリ１０６によってバックアップすることで、不揮発性メモリとし、主電源の切断時においても、バッファメモリ１０５の内容が約一週間保持されるため、データ喪失に至ることはない。ただし、信頼性の確保のために、通常の稼動において主電源を切断する際は、バッファメモリ１０５上のデータをすべてディスク装置１０９に反映させた後、主電源を切断する手続が採られる。

【００４５】ディスク装置１０９からホストにデータを読み出す場合は、前述のデータ書き込みの処理とは逆の処理が行われる。すなわち、ディスク装置１０９からバッファメモリ１０５上にデータを読み出し、当該バッファメモリ１０５に目的のデータがセーブされた時点でバッファメモリ１０５からホストにデータが転送される。この場合は、ディスク装置１０９に対するアクセスが先となるため、最初の読み出し処理では書き込みの場合のような早期応答が行えないが、たとえば、同一データに対する２回目以降の読み出しで、仮に、バッファメモリ１０５にデータが存在した場合、すなわちキャッシュヒットの場合には、ディスク装置１０９からの実際の読み出し動作を行うことなく、高速にアクセス可能なバッファメモリ１０５のデータを転送できるため、やはり、極めて高速なコマンドの実行が可能となる。

【００４６】すなわち、本実施例によれば、ホストからのＩ／Ｏ要求と、実際のディスク装置１０９に対するアクセスが分離され、ホストからのＩ／Ｏ要求は実質的に高速なバッファメモリ１０５との間で行われるので、ホストとの間で高スループットのデータ転送を行うことができる。

【００４７】次に、データ書き込み時にバッファメモリ１０５に空き領域が無くなった場合、およびデータの読み出し時にバッファメモリ１０５でミスヒットが発生し、ホストからのアクセス要求にディスク装置１０９のアクセスが伴う場合を考える。このような事象が発生するケースは、一般に複数のホストが磁気ディスクサブシステムを共有するマルチホスト構成で、極めて高頻度のＩ／Ｏ要求が発行される場合と考えられる。

【００４８】このような条件下では、単にディスク装置

１０９のシーク時間や回転待ち時間がコマンドの実行時間に含まれることによるコマンドの実行速度の低下の他に、多発するＩ／Ｏ要求の衝突によるホストインターフェイスＡ，ＢおよびドライブインターフェイスＣ，Ｄ（インターフェイスバス）の空き待ちによるスループットの低下が発生する。すなわち、バスネックによるスループットの低下および単位時間当たりのＩ／Ｏ処理数の低下である。この問題を解決するために、本実施例では、複数のドライブインターフェイスＣ，Ｄを設けるとともに、コントローラ部１００に、ディスク装置１０９がシーク中および回転待ち時にドライブインターフェイスＣ，Ｄを解放する制御論理、およびデータ転送時に任意の空いている側のドライブインターフェイスＣまたはＤを用いて動的に再接続する制御論理を持たせ、バスネックによるスループットの低下を回避している。

【００４９】図５に、このような制御論理による処理のタイムチャートの一例を示す。

【００５０】同図の例では、４台のホスト（Ｈｏｓｔ０，１，２，３）と４台のディスク装置１０９（Ｄｒｉｖｅ０，１，２，３）との間で、Ｈｏｓｔ０がＤｒｉｖｅ０に、Ｈｏｓｔ１がＤｒｉｖｅ１に、Ｈｏｓｔ２がＤｒｉｖｅ２に、Ｈｏｓｔ３がＤｒｉｖｅ３にリードコマンドを発行し、すべてミスヒットとなり、ディスク装置１０９に対するアクセスが必要となった場合を想定している。

【００５１】まず、区間５０１でＨｏｓｔ０からリードコマンドが発行されると、コントローラ部１００は区間５０２で一方のホストインターフェイスＡ（ＳＣＳＩ０）からコマンドを受領し、区間５０３でＤｒｉｖｅ０に対し、ドライブインターフェイスＣ（ＥＳＤＩ０）を介してシークコマンドを発行する。Ｄｒｉｖｅ０は、区間５０４でシークを実行する。このシーク中はＳＣＳＩ０およびＥＳＤＩ０は使われないため、区間５０５，区間５０６で当該ＥＳＤＩ０およびＳＣＳＩ０は解放される。

【００５２】引続き、区間５０７でＨｏｓｔ１からリードコマンドが発行されると、コントローラ部１００は、区間５０８で、空いているＳＣＳＩ０から当該コマンドを受理し、Ｄｒｉｖｅ１に対し区間５０９で空いているＥＳＤＩ０を介してシークコマンドを発行する。Ｄｒｉｖｅ１は区間５１０でシークを実行する。そして、Ｄｒｉｖｅ０の場合と同様に、区間５１１，区間５１２でＥＳＤＩ０およびＳＣＳＩ０は解放される。

【００５３】次に、Ｈｏｓｔ１によるリードコマンドの発行の区間５０７に僅かに遅れた区間５１３でＨｏｓｔ２からリードコマンドが発行される。

【００５４】このとき、ＳＣＳＩ０およびＥＳＤＩ０はＨｏｓｔ１からのコマンド発行に使用中であるため、ホストインターフェイスＢ（ＳＣＳＩ１）からコマンドが発行され、コントローラ部１００は区間５１４でコマンドを受理する。そして、目的のＤｒｉｖｅ２に対して区間５１５でドライブインターフェイスＤ（ＥＳＤＩ１）を介してシークコマンドを発行する。Ｄｒｉｖｅ２は、区間５１６でシークを実行し、さらに、区間５１７，区間５１８でＥＳＤＩ１およびＳＣＳＩ１は解放される。

【００５５】次に、区間５１９でＨｏｓｔ３からリードコマンドが発行される。このとき、ＳＣＳＩ０，ＥＳＤＩ０はすでに解放されているため、空いているＳＣＳＩ０から当該リードコマンドは発行され、コントローラ部１００は区間５２０で当該リードコマンドを受理する。そして、Ｄｒｉｖｅ３に対して区間５２１でＥＳＤＩ０を介してシークコマンドを発行する。Ｄｒｉｖｅ３は区間５２２でシークを実行する。このシーク中、区間５２３および区間５２４でＥＳＤＩ０およびＳＣＳＩ０は解放される。

【００５６】コマンド発行が終了後、各Ｄｒｉｖｅ０～３のシークが終了し、ヘッドが目的のセクタに達すると、各Ｄｒｉｖｅ０～３からコントローラ部１００に対して割り込みが発行され、これを契機に、コントローラ部１００は、Ｄｒｉｖｅ０～３およびＨｏｓｔ０～３の再接続を行い、データの転送を行う。

【００５７】図５の例では、最初にＤｒｉｖｅ０のシークが終了した場合を示している。

【００５８】区間５２５はＤｒｉｖｅ０の回転待ちを示し、区間５２６でＤｒｉｖｅ０からデータ転送が行われる。このデータは、区間５２７，区間５２８でＥＳＤＩ０およびＳＣＳＩ０を再接続し、区間５２９でＨｏｓｔ０に送られる。

【００５９】引続き、Ｄｒｉｖｅ１がシーク終了し、区間５３０での回転待ち時間後、当該Ｄｒｉｖｅ１は区間５３１でデータ転送を開始する。この時、ＥＳＤＩ０およびＳＣＳＩ０はＤｒｉｖｅ０からのデータ転送で占有されているため、空いている側のＥＳＤＩ１，ＳＣＳＩ１が区間５３２，区間５３３で再接続され、区間５３４でＨｏｓｔ１に送られる。

【００６０】Ｈｏｓｔ１からのリードコマンドは、ＳＣＳＩ０，ＥＳＤＩ０をから発行されていたが、再接続時には、他の空いている側のＳＣＳＩ１およびＥＳＤＩ１を選択することにより、従来のような固定的なインターフェイスの設定によるデータ転送におけるバスネックが解消される。

【００６１】なお、本実施例のコントローラ部１００では、このホストインターフェイスＡ，ＢおよびドライブインターフェイスＣ，Ｄの動的な切り離しおよび再接続を可能または不能に設定する制御論理を持っており、必要に応じて、データ転送中に使用する各インターフェイスを固定することも可能である。

【００６２】以下、同様に、Ｄｒｉｖｅ２，Ｄｒｉｖｅ３からの再接続を行い、全てのデータ転送が終了する。

【００６３】従来のように、Ｈｏｓｔ側およびＤｒｉｖ

e側に単一のインターフェイスしかなく、しかも当該各インターフェイスの動的な切り離しや再接続の機能がない場合、上述したＨｏｓｔ０～３とＤｒｉｖｅ０～３との間におけるデータ転送処理は全てシーケンシャルとなり、本実施例の場合に比較して、約2.５～３倍の処理時間を要することとなる。

【００６４】ここで、図５の区間５３５に注目すると、この間、ＳＣＳＩ０～１およびＥＳＤＩ０～１のすべてのインターフェイスにおいてDMA転送が行われるため、バッファメモリ１０５に対するアクセスが同時に進行することになる。

【００６５】そこで、これを同時処理するため、本実施例では、ホストインターフェイスA（ＳＣＳＩ０），ホストインターフェイスB（ＳＣＳＩ１）およびドライブインターフェイスC（ＥＳＤＩ０），ドライブインターフェイスD（ＥＳＤＩ１）の各々にＦＩＦＯメモリ１０２およびＦＩＦＯメモリ１０８を設け、DMA転送はサイクルスチールで時分割処理を行う。

【００６６】図６にDMA転送の経路周辺に着目した概念図を示す。

【００６７】ＦＩＦＯメモリ６０５（１０２），ＦＩＦＯメモリ６０６（１０２）の各バス６０１およびバス６０２は、それぞれＳＣＳＩ０，ＳＣＳＩ１に繋がり、ＦＩＦＯメモリ６０７（１０８），ＦＩＦＯメモリ６０８（１０８）の各バス６０３およびバス６０４は、ＥＳＤＩ０，ＥＳＤＩ１に繋がる。

【００６８】DMAコントローラ１０４は、ＦＩＦＯメモリ６０５～６０８のうちのいずれか一つを選択し、DMAバス１０３を介してバッファメモリ１０５との間でデータ転送制御を行う。バス６０１～６０４による転送を連続同時に行うためには、DMAバス１０３のデータ転送速度はバス６０１～６０４の全てのバス転送速度の合計を上回る必要がある。

【００６９】本実施例のコントローラ部１００では、Ｈｏｓｔ側に接続されるバス６０１，６０２は幅が１６ビットで最大１０MB／ｓｅｃであり、Ｄｒｉｖｅ側に接続されるバス６０３，６０４は幅が８ビットで最大5MB／ｓｅｃであり、DMAバス１０３は、３０MB／ｓｅｃ以上の転送速度が必要となる。このような高速転送を実現するため、本実施例のコントローラ部１００では、DMAバス１０３の幅を３２ビットとしており、当該DMAバス１０３とバス６０１，６０２およびバス６０３，６０４の幅変換を、ＦＩＦＯメモリ６０５～６０８で実現している。

【００７０】以上説明したように、本実施例の外部記憶装置によれば、ホスト側から見たアクセス要求に対する応答時間の大幅な短縮、さらには単位時間当たりのＩ／Ｏ要求発行件数を大幅に向上させることができるという効果が得られる。

【００７１】たとえば、本発明者らのシミュレーションによれば、本実施例で開示した構成の磁気ディスクサブシステムでは、バッファメモリを持たず、また単一のホストインターフェイスおよびドライブインターフェイスを持つ構成とした従来技術に比較して、平均発行Ｉ／Ｏ件数で６０％の向上、平均応答時間で約５７％の低減が期待できることが確認されている。

【００７２】以上本発明者によってなされた発明を実施例に基づき具体的に説明したが、本発明は前記実施例に限定されるものではなく、その要旨を逸脱しない範囲で種々変更可能であることはいうまでもない。

【００７３】

【発明の効果】本願において開示される発明のうち、代表的なものによって得られる効果を簡単に説明すれば、以下のとおりである。

【００７４】すなわち、本発明の外部記憶装置によれば、上位装置からの個々のリード／ライト要求に対する応答時間の短縮、および上位装置からのリード／ライト要求の発行件数の大幅な増加を実現することができるという効果が得られる。

【図面の簡単な説明】

【図１】本発明の一実施例である外部記憶装置の構成の一例を示すブロック図である。

【図２】本発明の一実施例である外部記憶装置の作用の一例を示す概念図である。

【図３】本発明の一実施例である外部記憶装置の作用の一例を示す概念図である。

【図４】本発明の一実施例である外部記憶装置の作用の一例を示すタイムチャートである。

【図５】本発明の一実施例である外部記憶装置の作用の一例を示すタイムチャートである。

【図６】本発明の一実施例である外部記憶装置において、DMA転送の経路に着目した概念図である。

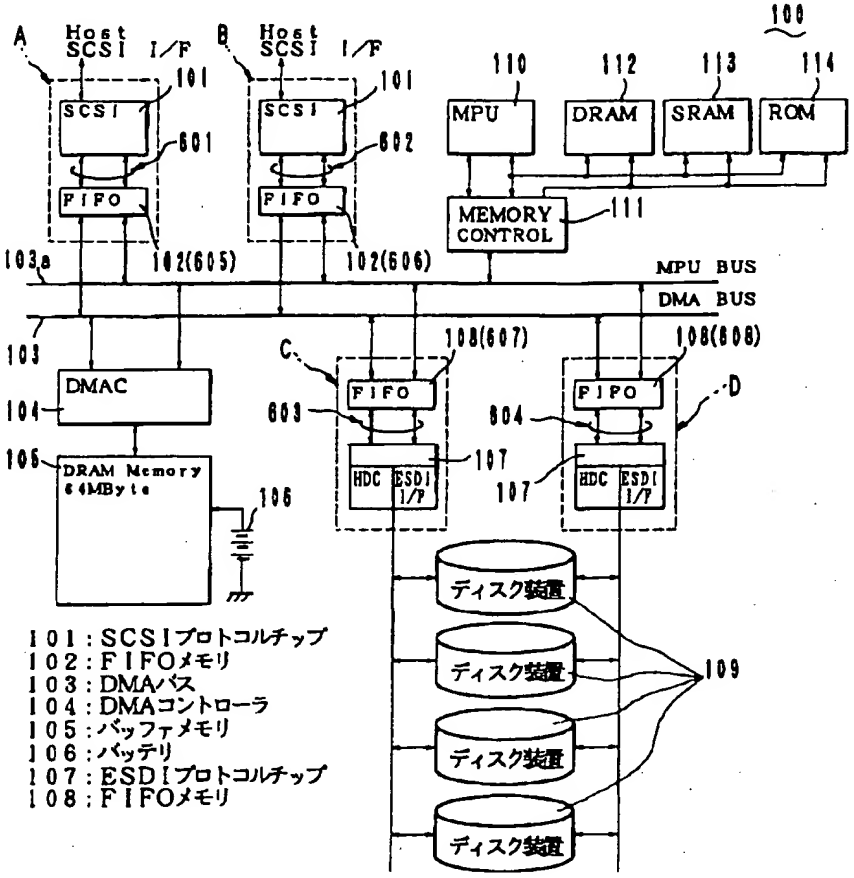
【符号の説明】

１００　コントローラ部（入出力制御部）
１０１　ＳＣＳＩプロトコルチップ
１０２　ＦＩＦＯメモリ（６０５～６０６）
１０３　DMAバス
１０３ａ　MPUバス
１０４　DMAコントローラ
１０５　バッファメモリ（一時データ保持手段）（第２の記憶媒体）
１０６　バッテリ
１０７　ＥＳＤＩプロトコルチップ
１０８　ＦＩＦＯメモリ（６０７～６０８）
１０９　ディスク装置（Ｄｒｉｖｅ０～３）（第１の記憶媒体）
１１０　マイクロプロセッサ
１１１　バスコントローラ
１１２　制御メモリ（DRAM）
１１３　制御メモリ（SRAM）

１１４　制御メモリ（ＲＯＭ）
２０１　割り込み信号
２０２～２０５　マイクロプロセッサ１１０の制御アクセス
２０６　バス
２０７　書き込みデータ
３０１～３０３　マイクロプロセッサ１１０の制御アクセス
３０４　バス
３０５　ディスク装置でのデータの格納位置
４０１　ホストからの書き込みデータ転送
４０２　ＳＣＳＩプロトコルチップからバッファメモリへのDMA転送
４０３　ディスク装置（Ｄｒｉｖｅ）でのシーク処理
４０４　回転待ち時間
４０５　バッファメモリからディスク装置（Ｄｒｉｖｅ）へのDMA転送
４０６　ディスク装置へのデータ転送
４０７　ホストへのコマンド終了報告（従来技術の場合）
４０８　コマンド実行時間（従来技術の場合）
４０９　ホストへのコマンド終了報告（本発明の場合）
４１０　コマンド実行時間（本発明の場合）
５０１　Ｈｏｓｔ０からＤｒｉｖｅ０へのリードコマンド発行
５０２　５０１のコマンド発行に対するＳＣＳＩ０の占有区間
５０３　５０１のコマンド発行に対するＥＳＤＩ０の占有区間
５０４　５０１のコマンド発行に対するＤｒｉｖｅ０のシーク
５０５　５０１のコマンド発行に対するＥＳＤＩ０の解放
５０６　５０１のコマンド発行に対するＳＣＳＩ０の解放
５０７　Ｈｏｓｔ１からＤｒｉｖｅ１へのリードコマンド発行
５０８　５０７のコマンド発行に対するＳＣＳＩ０の占有区間
５０９　５０７のコマンド発行に対するＥＳＤＩ０の占有区間
５１０　５０７のコマンド発行に対するＤｒｉｖｅ１のシーク
５１１　５０７のコマンド発行に対するＥＳＤＩ０の解放
５１２　５０７のコマンド発行に対するＳＣＳＩ０の解放
５１３　Ｈｏｓｔ２からＤｒｉｖｅ２へのリードコマンド発行
５１４　５１３のコマンド発行に対するＳＣＳＩ１の占有区間
５１５　５１３のコマンド発行に対するＥＳＤＩ１の占有区間
５１６　５１３のコマンド発行に対するＤｒｉｖｅ２のシーク
５１７　５１３のコマンド発行に対するＥＳＤＩ１の解放
５１８　５１３のコマンド発行に対するＳＣＳＩ１の解放
５１９　Ｈｏｓｔ３からＤｒｉｖｅ３へのリードコマンド発行
５２０　５１９のコマンド発行に対するＳＣＳＩ０の占有区間
５２１　５１９のコマンド発行に対するＥＳＤＩ０の占有区間
５２２　５１９のコマンド発行に対するＤｒｉｖｅ３のシーク
５２３　５１９のコマンド発行に対するＥＳＤＩ０の解放
５２４　５１９のコマンド発行に対するＳＣＳＩ０の解放
５２５　Ｄｒｉｖｅ０のシーク後の回転待ち時間
５２６　Ｄｒｉｖｅ０からのデータ読み出し
５２７　５２６のデータ読み出しに対するＥＳＤＩ０上でのデータ転送
５２８　５２６のデータ読み出しに対するＳＣＳＩ０上でのデータ転送
５２９　５２６のデータ読み出しに対するＨｏｓｔ０のデータ受理
５３０　Ｄｒｉｖｅ１のシーク後の回転待ち時間
５３１　Ｄｒｉｖｅ１からのデータ読み出し
５３２　５３１のデータ読み出しに対するＥＳＤＩ１上でのデータ転送
５３３　５３１のデータ読み出しに対するＳＣＳＩ１上でのデータ転送
５３４　５３１のデータ読み出しに対するＨｏｓｔ１のデータ受理
５３５　ＳＣＳＩ０～１およびＥＳＤＩ０～１のすべてでDMA転送が行われる区間
６０１～６０４　ＦＩＦＯメモリのバス
Ａ　ホストインターフェイス（ＳＣＳＩ０）（第１のインターフェイスポート）
Ｂ　ホストインターフェイス（ＳＣＳＩ１）（第１のインターフェイスポート）
Ｃ　ドライブインターフェイス（ＥＳＤＩ０）（第２のインターフェイスポート）
Ｄ　ドライブインターフェイス（ＥＳＤＩ１）（第２のインターフェイスポート）

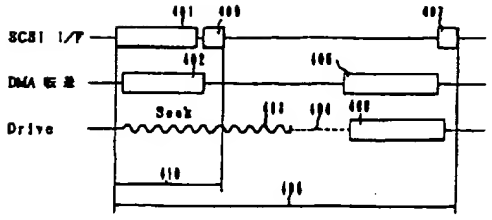
【図1】

図 1


100
A Host SCSI I/F
B Host SCSI I/F
101
SCSI
601
602
FIFO
MPU
DRAM
SRAM
ROM
110
112
113
114
MEMORY CONTROL
111
102(605)
102(606)
103a
MPU BUS
DMA BUS
103
C
D
108(607)
108(608)
DMAC
104
603
604
105
DRAM Memory 64MByte
106
HDC ESDI I/F
107
ディスク装置
109
101：SCSIプロトコルチップ
102：FIFOメモリ
103：DMAバス
104：DMAコントローラ
105：バッファメモリ
106：バッテリ
107：ESDIプロトコルチップ
108：FIFOメモリ

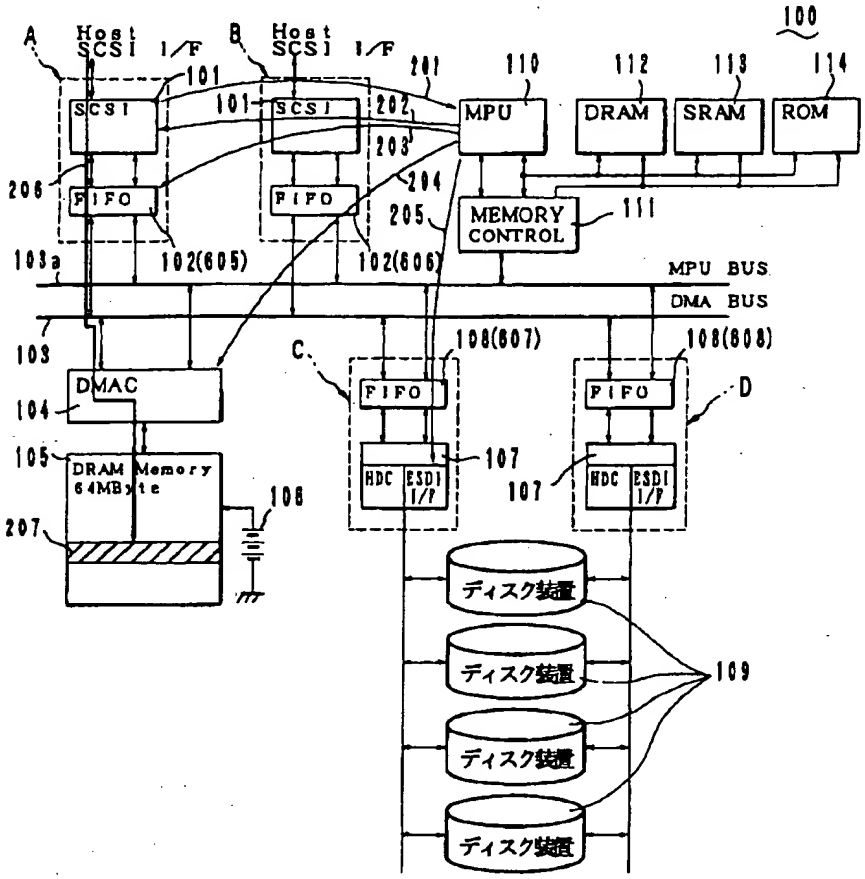
【図4】

図 4


SCSI I/F
DMA 転送
Drive
Seek

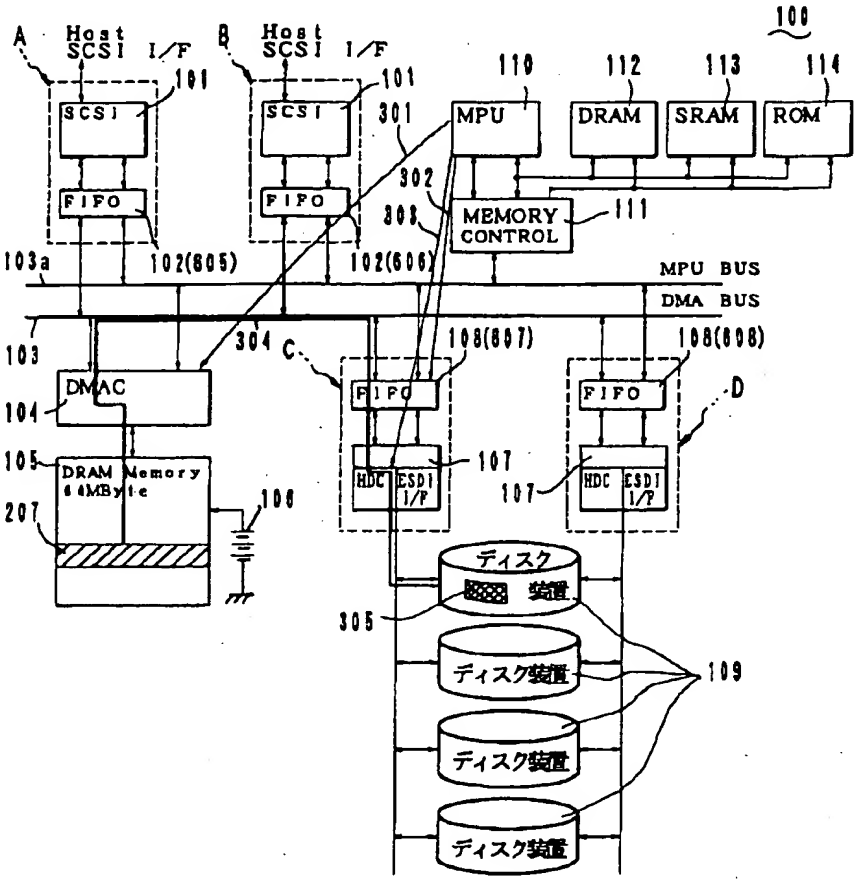
【図2】

図 2


100
A
Host SCSI I/F
B
Host SCSI I/F
101
101
SCSI
SCSI
201
202
203
204
205
206
FIFO
FIFO
MPU
110
DRAM
112
SRAM
113
ROM
114
MEMORY CONTROL
111
103a
102(605)
102(606)
MPU BUS
DMA BUS
103
DMAC
104
105
DRAM Memory 64MByte
106
207
C
108(607)
FIFO
HDC
SCSI I/F
107
D
108(608)
FIFO
HDC
SCSI I/F
107
ディスク装置
ディスク装置
ディスク装置
ディスク装置
109

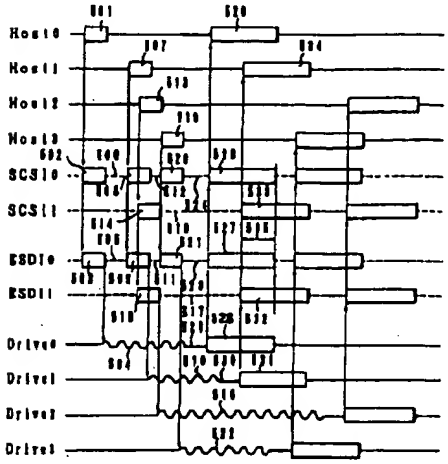
【図3】

図 3


100
A Host SCSI I/F
B Host SCSI I/F
101
SCSI
FIFO
102(605)
102(606)
110
MPU
112
DRAM
113
SRAM
114
ROM
301
302
303
MEMORY CONTROL
111
MPU BUS
DMA BUS
103a
103
304
C
D
108(607)
108(608)
DMAC
104
105
DRAM Memory 64MByte
106
207
HDC
ESDI I/F
107
ディスク装置
305
109

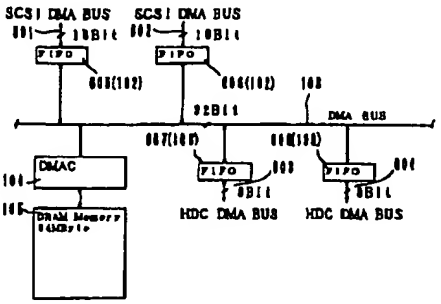
【図5】

図 5



【図6】

図 6



フロントページの続き

(72)発明者　中馬　顕
神奈川県小田原市国府津2880番地　株式会社日立製作所ストレージシステム事業部内

(72)発明者　湯川　芳雄
神奈川県小田原市国府津2880番地　株式会社日立製作所ストレージシステム事業部内